Machine Learning in the Energy Sector

[1]Maksym Oliinyk, [2]Jaroslav Džmura

[1]Department of Electric Power Engineering Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic
[2]Department of Electric Power Engineering Faculty of Electrical Engineering and Informatics Technical University of Košice, Slovak Republic

[1]maksym.oliinyk@tuke.sk, [2]jaroslav.dzmura@tuke.sk

***Abstract* — the so-called artificial intelligence (AI) is increasingly mentioned in the context of the technological transformation taking place at present. According to many experts, the use of artificial intelligence algorithms will affect almost all areas of the economy. In this regard, its impact on the energy sector seems interesting.**

***Keywords* — machine learning, smart grids, renewable energy, forecasting**

I. Machine Learning in the Energy Sector

An analysis of the publicly available information shows that the main areas of application of artificial intelligence in the energy sector at present are [1]:

- Forecasting tasks (meteorological information, equipment operating status, changes in consumption, etc.);
- Optimization tasks (operating modes of power system components, consumption, network configuration, etc.);
- Management tasks (artificial lighting, renewable energy and batteries, asset efficiency, etc.);
- tasks of communication (energy companies with consumers);
- Tasks of developing services and services (in terms of customer satisfaction with the range of services provided by companies, participation of enterprises in the work of energy markets, addressing quality assurance issues).

The expansion of the use of artificial intelligence tools in the energy sector will inevitably occur along with such unfolding processes as [2]:

- Energy transformation due to the increased use of local renewable energy sources, as well as the intellectualization of production, transmission and consumption of energy (smart technology);
- Digital transformation, due to the growing needs of monitoring and data analysis (big data) and the introduction of new technologies (for example, blockchain, "digital substation", unmanned devices for monitoring objects, etc.);
- The unification and mutual influence of various sectors of the energy and transport sectors (for example, Power-to-X technology). In light of the above, there is no doubt that the position of energy, as one of the most interesting areas of application of artificial intelligence methods will be strengthened.

Examples of the use of artificial intelligence algorithms in forecasting problems are enough today [3]. The dependence of renewable energy production on weather conditions has significantly increased the need for accurate forecasting. In the US state of Colorado, energy provider Xcel uses artificial intelligence algorithms to process information received from the National Center for Atmospheric Research (including satellite observation data in wind farm areas). This allows the company to generate detailed reports and optimize the operation of wind farms. IBM, together with the US Department of Energy (US Department of Energy), implements the SunShot initiative, in which a self-learning program can reliably predict the generation of renewable sources (solar, wind and hydro) [4]. The algorithm uses a large amount of historical data along with real-time weather monitoring information. According to analysts, the large-scale use of artificial intelligence algorithms to improve the operation of U.S. wind parks would theoretically allow them to increase their production in 2017 by 12 billion kWh and increase the share of wind energy in the total balance (6.3% last year). You can find examples of intellectualization tasks. PowerScout received two grants from the US Department of Energy for the development of cost-cutting programs for network companies and consumers (smart home), taking into account the integration of renewable sources [5]. Programs also use artificial intelligence algorithms. London Company Green Running Ltd. Develops a machine-learning-based Verv application designed to optimize the energy consumption of homes. The application works on computers, tablets and smartphones. The German company Schleswig-Holstein Netz AG, which operates electrical networks in the federal state of Schleswig-Holstein, reported an interesting application of artificial intelligence methods. Here, a self-learning network is used to identify the locations of the alleged damage. As the initial data, information is used on the life of the components of the electric networks and the repairs carried out, as well as information on loads and weather conditions. In addition, the American company AirFusion, which uses unmanned aerial vehicles to monitor the status of high-voltage power lines and wind power plants, uses software with artificial intelligence algorithms to process monitoring results. A neural network helps to better solve the problem of pattern recognition, for which, during the training process, thousands of images of damaged wind turbines are loaded into the program (including the consequences of lightning strikes, delamination, coating erosion, etc.) [6][7].

II. Regression models

Regression analysis is a method of modeling measured data and studying their properties. The data consists of pairs of values of the dependent variable (response variable) and the independent variable (explanatory variable). The regression model is a function of an independent variable and parameters with a random variable added. Model parameters are adjusted so that the model best approximates the data. The criterion for the quality of the approximation (objective function) is usually the standard error: the sum of the squares of the difference between the values of the model and the dependent variable for all values of the independent variable as an argument. Regression analysis is used to predict, analyze time series, test hypotheses, and identify hidden relationships in data [8].

The practice of regression analysis suggests that the linear regression equation often adequately expresses the relationship between the indicators even when in fact they turn out to be more complex. This is explained by the fact that, within the limits of the studied quantities, the most complex dependences can be approximately linear. In general form, the linear regression equation has the form [8]:

$$y = a_o + b_1 \cdot x_1 + b_2 \cdot x_2 + \ldots + b_k \cdot x_k + \varepsilon_i \qquad (1)$$

Where ***y*** - is a productive sign, the investigated variable; ***x*** - designation of the factor (independent variable); ***i*** is the total number of factors; and $\boldsymbol{a_0}$ is a constant (free) term of the equation; ***b*** - regression coefficient with a factor.

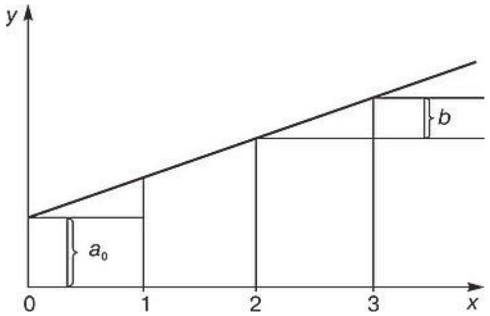

Fig. 1 Graph of a simple pairwise linear regression of $\boldsymbol{y = a_0 + b_x}$: the segment ***b*** shows the increment of ***y*** with increasing ***x*** by one

Phenomena are determined, as a rule, by a large number of cumulatively acting factors. In this regard, the problem often arises of studying the dependence of one variable ***y*** on several explanatory variables $\boldsymbol{x_1}$, $\boldsymbol{x_2}$ ... ***x***. This problem is solved using multiple regression analysis. The construction of the multiple regression equation begins with the solution of the question about the specification of the model, including the selection of factors and the choice of the type of regression equation. Factors included in multiple regression must meet the following requirements: they must be quantifiable (qualitative factors must be quantified); between factors there should not be a close correlation, and even less a functional dependence, i.e. there should be no multicollinearity. The inclusion of multicollinear factors in the model can lead to the following consequences: it is difficult to interpret the parameters of multiple regression as characteristics of the action of factors in a "pure form", since the factors are interconnected; linear regression parameters lose their economic meaning; parameter estimates are unreliable, have large standard errors and vary with the volume of observations [8].

Evaluation of the quality of the regression equation is carried out based on a set of criteria that verify the adequacy of the model to actual conditions and the statistical reliability of the regression. One of the most effective estimates of the adequacy of the model is the coefficient of determination of R^2. The true coefficient of determination of the model of the dependence of the random variable y on the factors x is determined as follows [9]:

$$R^2 = 1 - \frac{\sigma^2}{\sigma_y^2} \tag{2}$$

Where $\boldsymbol{\sigma_y^2}$ is the variance of the random variable **y**, $\boldsymbol{\sigma^2}$ — conditional (in terms of **x** factors) variance of the dependent variable (variance of model error).

R^2 describes the proportion of variation in the dependent variable due to regression or variability of the explanatory variables. The closer R^2 is to unity, the better the constructed regression model describes the relationship between the explanatory and the dependent variable. In the case of $R^2 = 1$, the studied relationship can be interpreted as functional (rather than statistical), which requires additional qualitative and quantitative information and changes in the research process [9].

III. Prediction of solar and wind energy production using machine learning

To demonstrate the operation of the regression models, three predicted variables were selected:

- The amount of energy that is generated by solar power plants in Germany;
- The amount of energy that is generated by wind power plants in Germany;
- The price of electricity in Germany.

Quite often in a single European energy system, critical situations arise related to the excessive production of electricity by renewable sources in Germany. This leads to violations of planned electricity supplies, cross-border lines overload, and price hikes in the electricity market. One possible solution to this problem is to more accurately predict the amount of energy generated by renewable energy sources. To build forecasting models, the Python programming language and the Sklearn machine-learning library were used [10]. This library allows you to quickly build regression models, as well as conduct a quick analysis of the results. To predict the used three regression models: linear, quadratic and cubic. The following data were used as input [11]:

1) Wind speed, km/h;
2) Solar radiation, W/m^2;
3) Direct solar radiation, W/m^2;
4) Air temperature, °C;
5) Installed power of solar power plants, MW;
6) Installed power of wind power plants, MW;
7) Electricity price, €/MWh;
8) Planned load power, MW.

For the first models and the simplicity of the graphical representation, only one attribute was used to train the model. With an increase in the number of features, the accuracy of model predictions increases, however, it is no longer possible to graphically present a model with six features. To predict the amount of energy in wind power plants, the sign of wind speed was used, and to predict the amount of energy in solar power plants, the sign of direct solar radiation was used. For models in which many features were used, only graphs of the overlay of the predicted results on the actual data are presented.

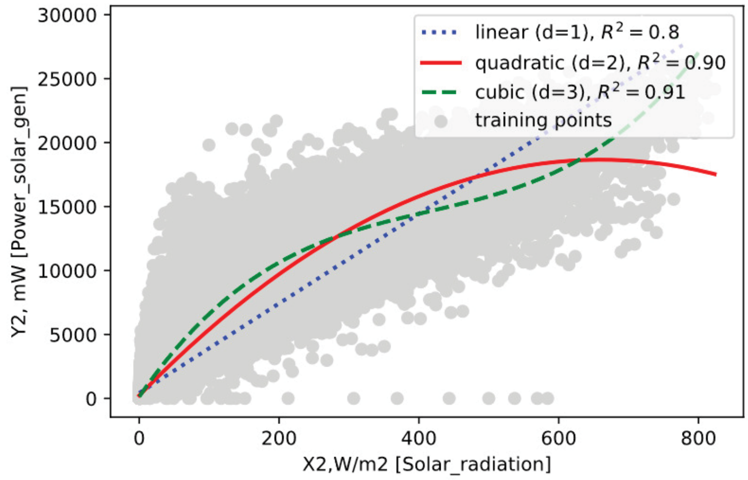

Fig. 2 Schedule of regression lines for generating energy from solar power plants.

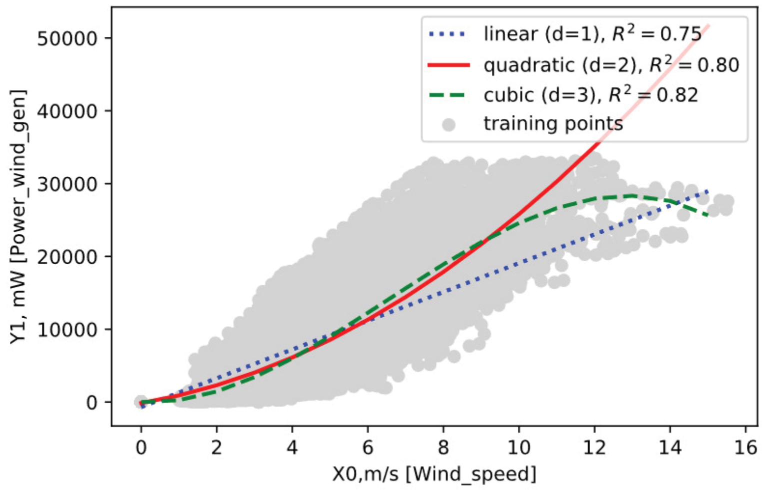

Fig. 3 Schedule of regression lines for generating energy from wind farms.

As can be seen in graphs 2 and 3 with the use of quadratic and cubic regression, the accuracy of predictions increases. However, a further increase in the degree of the polynomial does not give a significant increase, but it increases the operating time of the algorithm. To further increase the accuracy of forecasts, it is necessary to add additional features. In the future, quadratic regression for 6 features is used. The accuracy can be estimated by table 1, as well as on graphs 4, 5, 6.

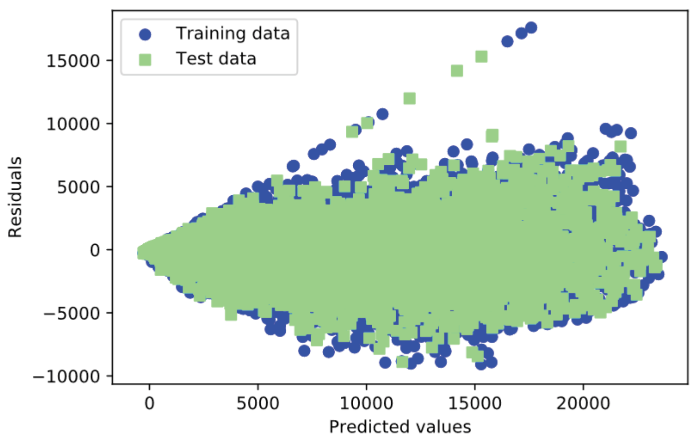

Fig. 4 Model predictive accuracy graph for solar power

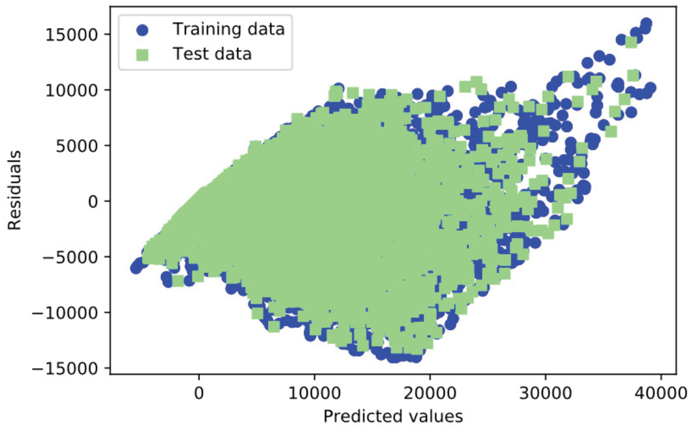

Fig. 5 Model predictive accuracy graph for wind power

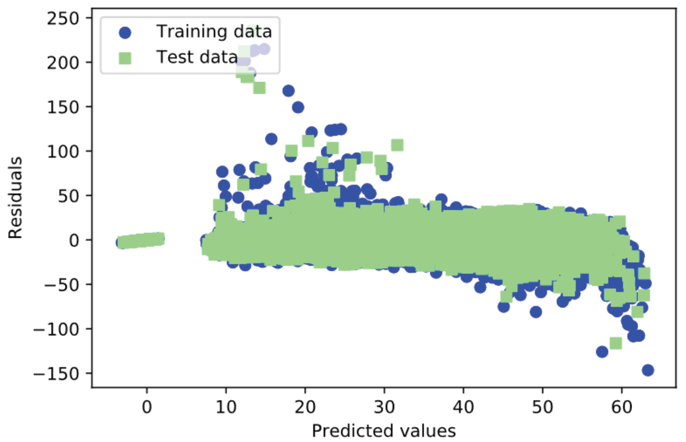

Fig. 6 Model predictive accuracy graph for electricity prices

As can be seen in the fig. 4, 5, 6, the accuracy of the forecasts is quite high. On the graphs, you can see that the worst of the entire model is able to predict values with a strong deviation. To predict such values, a large sample is needed to train the model, as well as additional features, such as wind direction, data on local hurricanes (storms), and repairs on power lines / substations / power plants. You can see the data on the indicator (coefficient of determination of $R^{2)}$ R^2 in table 1. The closer the values of this indicator are to 1, the higher the accuracy of the model predictions.

Table 1
Coefficient of Determination of R^2

Model	R^2
*1 Feature predictions **solar power** (linear)*	0.8
*1 Feature predictions **solar power** (quadratic)*	0.9
*1 Feature predictions **solar power** (cubic)*	0.91
*1 Feature predictions **wind power** (linear)*	0.75
*1 Feature predictions **wind power** (quadratic)*	0.80
*1 Feature predictions **wind power** (cubic)*	0.82
*6 Features predictions **wind power***	0.87
*6 Features predictions **solar power***	0.94
*6 Features predictions **price electricity***	0.79

IV. Conclusion

Accuracy of the model predictions is quite high, especially in relation to the energy generated by solar power plants; this is primarily due to the constancy of solar radiation. To improve the forecasting of the amount of energy in wind power plants, it is necessary to add such a feature as the direction of the wind, this parameter plays a rather important role in the possibility of energy production at this type of power plant. The accuracy of forecasting electricity prices is not the highest. This is mainly because not only Germany but also a number of neighboring countries have access to the exchange (where the price is set), to improve the accuracy of forecasts, we need data on energy consumption, as well as energy generated on renewable energy in these countries.

With the increase in the number of renewable energy sources, as well as the number of smart meters, the relevance of machine learning in the electric power industry will only increase.

ACKNOWLEDGMENT

This work was supported by the Scientific Grant Agency of the Ministry of Education of Slovak Republic and the Slovak Academy of Sciences by the projects VEGA No. 1/0372/18

REFERENCES

[1] A. Mogilenko, "Application of artificial intelligence algorithms in the global energy industry", in *Energy and Industry of Russia*, 2018, no 7, p. 23.

[2] A. Losev, "Artificial Intelligence in Nuclear Power", in *Atomic Expert*, 2018, no 3-4, pp. 35-43.

[3] E. Kamchatova, A. Vasilieva, N. Lyasnikov, M. Dudin, N. Vysotskaya, "Energy saving management in urban economy and industry" in *International Journal of Civil Engineering and Technology*, 2018, vol. 9, no 6, pp. 1423-1429.

[4] American Wind Energy Association, "U.S. wind industry annual market report year ending 2012," Amer. Wind Energy Assoc., Washington, DC, USA, 2013 [Online]. Available: http://awea.files.cmsplus.com/images/AWEA_USWindIndustryAnnualMarketReport2012_ ExecutiveSummary(2).pdf.

[5] "How does the domestic business connect to digitalization projects in the power industry", 24 RosInfo, (Unpublished work style)," unpublished

[6] R. Jursa and K. Rohrig, "Short-term wind power forecasting using evolutionary algorithms for the automated specification of artificial intelligence models", in *Int. J. Forecast*., vol. 24, pp. 694–709, 2008.

[7] M. Monfared, H. Rastegar, and H.M. Kojabadi, "A new strategy for wind speed forecasting using artificial intelligent methods", *Renew. Energy*, vol. 34, pp. 845–848, 2009

[8] D. A. Freedman, *Statistical Models: Theory and Practice,* Cambridge University Press. p. 26.

[9] Warne, T. Russell (2011). "Beyond multiple regression: Using commonality analysis to better understand R2 results". Gifted Child Quarterly. 55 (4): 313–318. doi:10.1177/0016986211422217

[10] Scikit-learn: Machine Learning in Python, Pedregosa *et al.*, JMLR 12, pp. 2825-2830, 2011.

[11] Open Power System Data. [Online]. Available: https://data.open-power-system-data.org